ADENOSINE-RICH SEQUENCES IN RAPIDLY HYBRIDIZING MESSENGER-LIKE RNA AND THEIR POSSIBLE SIGNIFICANCE FOR REITERATED BASE SEQUENCES IN EUKARYOTIC DNA

Klaus SCHERRER
Swiss Institute for Experimental Cancer Research, Lausanne, Switzerland

Received 24 June 1971

Several years ago we demonstrated by hybridization to homologous DNA the occurrence of adenosine-rich (A-rich) sequences in the rapidly labelled mlRNA [1, 2] from avian erythroblasts [3].

We interpreted this observation as demonstrating the existence in mlRNA, and particularly in the nascent fraction, of homologous signal or spacer loci, distributed over the genome [2]. Recently the existence of A-rich sequences in vaccinia-virus specific mRNA [4] and in HeLa cell RNA was described [5] and A-rich sequences were found in globin mRNA [6].

The discussion concerning repetitive sequences in eukaryotic DNA [7–9], their possible role in gene regulation [10, 11] and their practical importance in molecular hybridization with mlRNA lead us to re-examine the possible implications of the occurrence of A-rich sequences in the eukaryotic genome.

In the following report we will describe and discuss in detail the original observations, mentioned previously as a peripheral property of mlRNA [3, 12] We shall also discuss recent confirmatory evidence on such sequences obtained using HeLa cell messenger and mlRNA.

Abbreviations:

A	: adenosine
U	: uridine
G	: guanosine
C	: cytidine-monophosphates
mRNA	: messenger RNA
mlRNA	: messenger-like RNA [1, 2]
SSC	: standard saline citrate
TLC	: thin layer chromatography
A	: absorbance
UV	: ultraviolet light

2. Material and methods

Duck erythroblasts and HeLa cells were labelled with ^{32}P or ^{3}H-uridine and RNA was extracted, purified and fractionated as previously described [3, 13, 19]. Hybridization was carried out either in liquid phase and evaluated by the Sephadex technique [3] or on filters by the urea low-temperature technique [2]. In the latter case RNA was isolated from nitrocellulose filters either by heating for 15 min to 90° with 0.01 × SSC or to 60° in the same buffer with 50% formamide.

Base compositions were determined either chromatographically [3] or electrophoretically [14] on thin layer cellulose (TLC) plates after alkaline hydrolysis. Such determinations were always performed at least twice.

3. Results

Table 1 summarizes the results of a series of experiments carried out with RNA from duck erythroblasts. The base composition of pulse labelled mRNA and mlRNA is characterized by a high content of U

Table 1
Base composition of hybridized and control erythroblast RNA.

Determ. no.	RNA fraction	Labelling time (min)	Hybridized	Method	Base composition (%)			
					A	U	G	C
1	total	45	no	UV-abs.	20.3	18.7	30.5	30.5
2	total	45	yes	UV-abs.	20.9	*27.9*	25.4	25.7
3	total	15	no	^{32}P	21.6	27.3	28.8	22.2
4	total	15	yes	^{32}P	*31.3*	26.8	31.3	10.4
5	total	45	no	^{32}P	21.7	26.6	27.0	24.5
6	total	45	yes	^{32}P	*47.1*	15.5	28.8	8.6
7	total	120	no	^{32}P	21.7	23.9	26.4	27.9
8	total	120	yes	^{32}P	*60.2*	10.6	17.6	11.5
9	45 – 70 S	15	no	^{32}P	22.4	27.8	25.7	23.9
10	45 – 70 S	15	yes	^{32}P	30.4	23.9	28.2	17.4
11	30 – 45 S	15	no	^{32}P	22.5	28.0	25.4	24.1
12	30 – 45 S	15	yes	^{32}P	30.5	27.7	23.8	17.8
13	45 – 70 S	45	yes	^{32}P	*35.4*	28.1	20.9	15.4
14	30 – 45 S	45	yes	^{32}P	*39.0*	19.5	27.0	14.4
15	10 – 30 S	45	yes	^{32}P	*43.9*	19.7	23.0	13.2
16	28 S rRNA	–	–	UV-abs.	20.5	19.3	28.3	31.9
17	DNA	–	–	UV-abs.	28.5	28.1	21.2	22.3

Duck erythroblasts were labelled for the indicated times with carrier free ^{32}P (10 mCi/ml cells). RNA was extracted, purified and fractionated on sucrose gradients. Hybridization: 500 μg DNA, 4000 μg RNA, 10 ml 2 × SSC, 65°, 4 days, during last 12 hr temperature gradually lowered to 30°. RNase (pancreatic, no RNase T_1) treatment: 30 μg/ml, 25°, 3 hr. Hybrid isolated in the exclusion volume on Sephadex G-200, precipitated with TCA, hydrolysed with 0.3 M KOH. Base compositions were determined on TLC plates as indicated in Material and methods. For determ. 2, individual spots after 2-dimensional chromatography were removed from the plates and the nucleotides eluted with 0.1 N HCl and purified by quantitative adsorption/desorption on charcoal. Concentrations of nucleotides were determined spectrophotometrically, controlling the characteristic spectrum for the individual nucleotides.

[15–18], as shown best in determin. 9 (15 min pulse, RNA with S > 45). In erythroblasts, ribosomal RNA synthesis is normally low, but cannot be arrested completely with low doses of actinomycin (which interferes with erythropoiesis at very low doses). Therefore during prolonged exposure to ^{32}P, increasing amounts of labelled pre-rRNA and some rRNA shift the overall base composition to higher C and lower U contents, A or G remaining fairly constant (table 1, determ. 3, 5 and 7).

The base composition in the hybrid formed by incubating total cell RNA with homologous DNA (4 days, 65°, 2 × SSC) is essentially the same (28% U and 21% A) as that of the mlRNA (input mlRNA) if determined spectrophotometrically. By measuring actual concentrations of nucleotides we can determine the composition of the hybridized RNA which corresponds to the steady state population of mlRNa, i.e., the pool of metabolized and stabilized molecules in nucleus or cytoplasm [2]. The base composition of the hybrid, identical to that of the steady state mlRNA suggests that RNA corresponding to average structural genes has been hybridized to some extent during the 4 days of incubation. The fraction of DNA engaged in this hybrid amounts to 4.5% [1]. This demonstrates that a large spectrum of DNA sequences must have participated in the reaction.

However, the RNA base composition, if determined by ^{32}P distribution in the same hybrid

Table 2
Base composition of hybridized and control HeLa cell RNA.

Determ. no.	RNA fraction	Hybridization		RNase treated	Base composition (%)			
		done	time (days)		A	U	G	C
1	mlRNA, nuclear (S > 60)	no	–	–	19.4	31.3	27.9	20.4
2	mlRNA, nuclear (10 – 150 S)	yes	7	no	22.7	29.2	22.7	25.4
3	mlRNA, nuclear (10 – 150 S)	yes	7	yes	22.2	26.7	25.8	25.3
4	mlRNA, nuclear (10 – 150 S)	yes	21	no	24.1	30.8	22.9	23.6
5	mlRNA, nuclear (10 – 150 S)	yes	21	yes	22.4	27.4	30.9	19.2
6	mlRNA, cytoplasm (6 – 30 S)	no	–	–	24.9	27.8	21.6	25.6
7	mlRNA, cytoplasm (6 – 30 S)	yes	21	yes	*28.7*	26.9	21.5	23.9
8	mRNA, polysomal (6 – 30 S)	no	–	–	22.8	27.0	24.1	25.8
9	mRNA, polysomal (6 – 30 S)	yes	21	yes	*27.6*	26.3	23.8	23.1

HeLa cells were labelled for 6 hr with carrier free ^{32}P (50 mCi, 2 l at 4.5 × 10^5 cells/ml; 0.01 times normal phosphate conc., 100 μg/ml toyocamycine (to inhibit rRNA synthesis). RNA was extracted and purified according to standard methods. One part was fractionated by acrylamide gel electrophoresis according to Mirault and Scherrer [25]. Hybridization was carried out according to Material and methods for the times indicated: 20 μg DNA/filter, 2 ml 4 × SSC, 7 M urea, 41°. After washing free of RNA with 2 × SSC, filters were incubated for 24 hr in 10 ml 4 × SSC, 7 M urea, before pancreatic RNase (20 μg/ml, 1 hr, 25°). Base compositions were determined according to Material and methods.

(determ. 2 and 6) gives a strikingly different result: the A content is higher than that of U. It rises from 31% (15 min labelling, determ. 4) to as much as 60%, if the input RNA was labelled for 2 hr. At the same time the values of U and G fall. The G content serves as an internal control for the absence of contamination by an RNase resistant core, rich in purines. (No T_1 RNase was added to the pancreatic RNase in order to avoid the possible selection among purine-rich sequences of the core for the A-rich fraction.)

Another clear correlation, less dramatic than that of the increase of A in the hybrid with increased labelling time of the input RNA, is that the smaller the RNA molecules the higher the A content in the hybrid, as shown by determ. 13–15 in table 1.

Both correlations lead to the same conclusion: small and metabolised mlRNA molecules have relatively more A-rich sequences than the giant nascent precursor mlRNA.

We took advantage of the possibility to label mlRNA in HeLa cells to constant specific activity [19] in order to check the base composition of RNA hybridizing within different periods of time. The results concerning short hybridization (6 to 24 hr), show in general an increase in A content of the hybrid relative to that of the mlRNA. However, since only low activities could be obtained, the validity of these observations is uncertain at present. In table 2 (determ. 1–5) are the results for long term hybridization (7 and 21days). The base composition of the hybrid RNA is relatively close to that of the input mlRNA. This confirms the result obtained with unlabelled erythroblast RNA (table 1, determ. 2). No selection for A-rich sequences occurs if the RNA is labelled to steady state, and if the hybridization reaction has reached equilibrium. Surprising, however, is the finding that hybrids formed with cytoplasmic mlRNA and polyribosomal mRNA show a distinct increase in A (table 2, determ. 6–9), even in long term hybridization.

4. Discussion

The main observation in these experiments is the demonstration of A-rich sequences in mlRNA by hybridization. The fact that the content of adenosine is not particularly high in mlRNA before hybridization, and the paradoxical difference in base composition of the hybrid RNA, observed if the optical measurement of ribonucleotide concentrations (UV-base composi-

tion) is compared to the ^{32}P distribution, suggests that particular sequences are selected during the DNA–RNA interaction.

We think that A-rich sequences may have a two fold advantage in hybridization: firstly because of rapidity of interaction due to the homology of the sequence and secondly as a consequence of the existence of a great number of T-rich sequences in DNA relative to the number of transcribed informational sequences.

In fact, the apparent paradox, the difference of UV and ^{32}P base composition in the hybrid, may be explained by assuming that every mlRNA molecule contains one or several relatively short A-rich sequences in covalent linkage with the informational (message) and, possibly, with no-sense (spacer) areas (comparable to the rejected parts of pre-rRNA). Such an RNA structure would mean that T-rich sequences exist *all over the genome,* not only in the fraction transcribed into RNA but also in the temporarily silent part of DNA. Thus, the A-rich part of mlRNA could hybridize all over the genome whereas the informational (or spacer) part of the same molecules would be restricted in hybridization to the actually transcribed DNA. Due to their relatively high frequency rapidly successful interactions between these homologous A- respectively T-rich sequences can be expected. Thus, an mlRNA molecule could be annealed at a site of the DNA (in transcribed or silent regions) where the message (or spacer) part of the same molecule finds no sequence homology. As a consequence, no, or only imperfect hydrogen bonding would be possible between the DNA and this non-complementary RNA sequence. During the selection process imposed by pancreatic RNase treatment after hybridization these pseudo-hybrids would be eleminated leaving intact the A-T homologous part.

In DNA–DNA "renaturation" however, where this selection by an enzyme which leaves only true double stranded regions intact is not feasible, the pseudobase-paired regions would account for double stranded DNA, leading to a gross overestimation of its amount.

The A-rich regions of the short-labelled erythroblast mlRNA can be observed by ^{32}P distribution and not in the UV – base composition since the informational molecules (mRNA and stabilized mlRNA) are largely unlabelled. The transcribed DNA contains larger areas homologous to the informational (or spacer) part of mlRNA than to A-rich sequences as demonstrated by the actual base composition of mlRNA. Thus, informational (or spacer) sequences will represent the bulk of the hybridized RNA. However, since a large fraction of the nascent RNA is turning over [2], the specific activity of the A-rich sequences among the hybridized RNA molecules will be higher than that of the informational (or spacer) part, unless the RNA is labelled uniformly.

This conclusion is corroborated by the fact that the RNA hybrid formed by HeLa mlRNA labelled with ^{32}P to constant specific activity [19] has a base composition only slightly higher in A than that of the input RNA (22%, compared to 19%), if the hybridization is brought to equilibrium (7–21 days). The hybrid of pure mRNA, however, shows a significant enrichment in A. This may be attributed to relatively larger A-rich sequences in cytoplasmic mlRNA and mRNA, compared to nuclear mlRNA. Since nuclear mlRNA hybridizes to about 5% of the DNA, mRNA with about ten times less DNA [1, 2], the A-rich sequences hybridizing all over the genome have a 10 times higher statistical chance to be selected in the case of mRNA.

Preliminary experiments show that in short term hybridization, where an advantage is given to highly reiterated sequences (6 hr to 24 hr) the base composition of the hybrid RNA before the RNase treatment is also equal to that of the input RNA. An increase in A content – less dramatic than in the case of the pulse labelled erythroblast RNA – is observed only after the destruction of the pseudo-hybrid by RNAse. It is particularly pronounced in case of polyribosomal mRNA.

An alternative explanation for the increase in A-rich sequences in the hybrid with increased labelling time and after cleavage to small sized RNA (table 1, determ. 4, 6, 8 and 13, 14, 15 resp.) may be the further terminal addition of A during processing of mlRNA. Thus, the poly A stretch in 3′ position of globin [6] mRNA might be added after the transcription of the globin sequence.

We conclude that mlRNA and mRNA must contain A-rich sequences in covalent linkage with informational (or spacer) sequences. They may account for the very rapidly renaturing sequences in DNA–DNA and

DNA–RNA interactions. They could be at the basis of the rapid hybridization of 5′ and 3′ ends of mlRNA [20, 21].

It may be speculated that their counter-parts in the DNA provide attachment sites for RNA-polymerase or regulatory proteins. It is interesting to note, that the (silent) satellite DNA in the mouse is particularly rich in T isostiches [22] and is localized in centrometric chromosome areas. Possibly the "interchromomere" DNA, linking transcribed areas (transcriptional units; puffs in giant chromosomes) is composed of T-rich sequences: the RNA polymerase could start to be transcribed from these areas into the transcriptional unit containing the message.

The occurrence of A-rich sequences in mRNA may also provide attachment sites for the specific transport or regulatory proteins associated with globin mRNA observed in erythroblasts of avian [23] or mammalian sources [24].

Acknowledgements

We thank Mrs. L. Richardson-Voegelin and Miss Rita Mueller for skilful technical assistance. This work was supported by the French Centre National de la Recherche Scientifique and by the Swiss National Foundation.

References

[1] K. Scherrer and L. Marcaud, J. Cell Physiol. 72, Suppl. 1 (1968) 181.
[2] K. Scherrer, G. Spohr, N. Granboulan, C. Morel, J. Grosclaude and C. Chezzi, Cold Spring Harbor Symp. Quant. Biol. 35 (1970) 539.
[3] K. Scherrer, L. Marcaud, F. Zajdela, B. Breckenridge and F. Gros, Bull. Soc. Chim. Biol. 48 (1966) 1037.
[4] J. Kates, Cold Spring Harbor Symp. Quant. Biol. 35 (1970) 743.
[5] M. Edmonds and M.H. Vaughan, Jr., Federation Proc. 29 (1970) 2397.
[6] L. Lim and E.S. Canellakis, Nature 227 (1970) 740.
[7] R.J. Britten and D.E. Kohne, Science 161 (1968) 529.
[8] P.M.B. Walker, Nature 219 (1968) 228.
[9] B. McCarthy, Biochem. Genet. 2 (1968) 37.
[10] R.J. Britten and E.H. Davidson, Science 165 (1969) 349.
[11] G.P. Georgiev, J. Theoret. Biol. 25 (1969) 473.
[12] K. Scherrer, personal communication, Gordon Conf. (1966).
[13] K. Scherrer, Fundamental Techniques in Virology, ed. K. Habel and N.P. Salzman (Academic Press, New York, 1969) p. 413.
[14] A. Hiby and B. Kröger, J. Chromat. 26 (1967) 545.
[15] S. Penman, K. Scherrer, Y. Becker and J.E. Darnell, Proc Natl. Acad. Sci. U.S. 49 (1963) 651.
[16] K. Scherrer and L. Marcaud, Bull. Soc. Chim. Biol. 47 (1965) 1697.
[17] K. Scherrer, L. Marcaud, F. Zajdela, I.M. London and F. Gros, Proc. Natl. Acad. Sci. U.S. 56 (1966) 1571.
[18] R. Soiero, H.C. Birnboim and J.E. Darnell, J. Mol. Biol. 19 (1966) 362.
[19] G. Spohr, N. Granboulan, C. Morel and K. Scherrer, European J. Biochem. 17 (1970) 296.
[20] Ch. Coutelle, A.P. Ryskov and G.P. Georgiev, FEBS Letters 12 (1970) 21.
[21] A.P. Ryskov, V.L. Mantieva, A.R. Avakain and G.P. Geogiev, FEBS Letters 12 (1971) 141.
[22] E.M. Southern, Nature 227 (1970) 794.
[23] C. Morel, B. Kayibandan and K. Scherrer, Experientia 27 (1971) 22.
[24] J. Marbaix, personal communication (1971).
[25] M.-E. Mirault and K. Scherrer, European J. Biochem., submitted.